(1)

CONTRACT NO.: DAMD17-88-C-8126

TITLE: CLONING AND BIOCHEMICAL CHARACTERIZATION OF HIV INTEGRASE

PRINCIPAL INVESTIGATOR: ELLEN MURPHY

PI ADDRESS: The Public Health Research Institute
455 First Avenue
New York, NY 10016

REPORT DATE: December 15, 1989

TYPE OF REPORT: Midterm

PREPARED FOR: U.S. ARMY MEDICAL RESEARCH AND DEVELOPMENT COMMAND
FORT DETRICK
FREDERICK, MARYLAND 21701-5012

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

1a. REPORT SECURITY CLASSIFICATION Unclassified	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution unlimited
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S)	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION The Public Health Research Institute 6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION
6c. ADDRESS (City, State, and ZIP Code) New York, NY 10016	7b. ADDRESS (City, State, and ZIP Code)
8a. NAME OF FUNDING/SPONSORING ORGANIZATION U.S. Army Medical Research & Development Command 8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER DAMD17-88-C-8126
8c. ADDRESS (City, State, and ZIP Code) Fort Detrick Frederick, MD 21701-5012	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO. 63105A; PROJECT NO. 3M2-63105DH29; TASK NO. AC; WORK UNIT ACCESSION NO. 088

11. TITLE (Include Security Classification)
(U) Cloning and Biochemical Characterization of the HIV Integrase

12. PERSONAL AUTHOR(S)
Ellen Murphy

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Midterm	FROM 6/6/88 TO 12/15/89	1989 December 15	

16. SUPPLEMENTARY NOTATION

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	
06	03		RA 1; Viral DNA; Viral endonuclease; Retrovirus; Lentivirus; Genome; Molecular virology (KT)
06	13		

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

Our long range goal is to understand the macromolecular interactions responsble for the integration of HIV DNA into human chromosomes. The primary goal of this project is to develop appropriate biochemical assays for the interaction of the HIV integration protein, IN, with a specific DNA target, the viral LTRs. The gene encoding IN has been subcloned from infectious viral DNA into appropriate cloning vectors and has been expressed in both *E. coli* and in insect cells. IN has been expresed as a *pol* fusion protein that is processed by HIV protease *in vivo* and as a single protein carrying an added methionine residue. The lack of any additional carboxy-terminal processing of IN has beer demonstrated. Purification of the recombinant IN is in progress. Potential DNA substrates for IN have been constructed. These include a synthetic oligonucleotide substrate corresponding to the LTR end and a substrate containing the ligated junction of the two LTRs, obtained by polymerase chain amplification from HIV-infected cells. The latter provides both linear and circular DNA forms. During the construction of the circle junction clone, it was found that 50% of the circle junction sequences in infected cells are aberrant, containing deletions or insertions at the circle junction. The involvment of IN in the production of these forms will be investigated. (Keywords:)

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT	21. ABSTRACT SECURITY CLASSIFICATION
☐ UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT. ☐ DTIC USERS	Unclassified
22a. NAME OF RESPONSIBLE INDIVIDUAL Mary Frances Bostian	22b. TELEPHONE (Include Area Code) 301-663-7325
	22c. OFFICE SYMBOL SGRD-RMI-S

DD Form 1473, JUN 86 Previous editions are obsolete. SECURITY CLASSIFICATION OF THIS PAGE

FOREWORD

Opinions, interpretations, conclusions and recommendations are those of the author and are not necessarily endorsed by the U.S. Army.

_____ Where copyrighted material is quoted, permission has been obtained to use such material.

_____ Where material from documents designated for limited distribution is quoted, permission has been obtained to use the material.

_____ Citations of commercial organizations and trade names in this report do not constitute an official Department of the Army endorsement or approval of the products or services of these organizations.

EM In conducting research using animals, the investigator(s) adhered to the "Guide for the Care and Use of Laboratory Animals," prepared by the Committee on Care and Use of Laboratory Animals of the Institute of Laboratory Animal Resources, National Research Council (NIH Publication No. 86-23, Revised 1985).

_____ For the protection of human subjects, the investigator(s) have adhered to policies of applicable Federal Law 45CFR46.

EM In conducting research utilizing recombinant DNA technology, the investigator(s) adhered to current guidelines promulgated by the National Institutes of Health.

Ellen Murphy 12/15/89

PI Signature Date

Accesion For	
NTIS CRA&I	☑
DTIC TAB	☐
Unannounced	☐
Justification	
By	
Distribution /	
Availability Codes	
Dist	Avail and / or Special
A-1	

TABLE OF CONTENTS

FIGURES

INTRODUCTION

Our long range goal is to understand the macromolecular interactions responsible for the integration of HIV DNA into human chromosomes with the specific purpose of identifying compounds that block integration. Since integration is an early step in the retroviral life cycle, blocking integration prevents cells from becoming infected. Moreover, integration is not a part of normal cell physiology; consequently, compounds that block integration should be good candidates as therapeutic agents for AIDS. What follows in this section is a brief review of relevant biochemical aspects of retroviral integration and an outline of our experimental rationale.

Genetic analyses of several retroviral systems have identified three regions of the viral genome essential for the integration reaction and for productive infection. One is the 3′ end of *pol*, a gene called *int* (7, 27, 28, 30). In the case of avian retroviruses the product of *int*, the integrase protein (IN), exhibits a nuclease activity with preference for nucleotide sequences expected to be involved in integration (5, 8, 13-15, 21, 25). The other two regions are the putative viral targets of IN, the ends of viral DNA (5, 6, 26).

One of our aims is to develop biochemical assays for integration. For this our first task is to define the substrates of the reaction. Although the genetic analyses cited above clearly implicated IN and the viral ends as components in the reaction, the topological form of the viral DNA substrate was not initially obvious because cells contain three forms. Early *in vivo* studies with an avian virus had indicated that circular viral DNA containing two tandem LTRs (2-LTR circles) is capable of integration (27). However, recent reports using a murine system indicate that this is not a general phenomenon (9, 22). Moreover, IN is not thought to produce the correct type of cleavage with a circle. A simple staggered endonucleolytic cleavage at the junction between the two LTR's (the circle junction) in a 2-LTR circle would not produce a form of linear viral DNA compatable with the integration reaction intermediates that have been found (4, 11). Thus circles are now believed to be off the main integration pathway.

As *in vitro* integration assays were developed for MuLV, a reaction intermediate was identified in which the 3′ ends of the viral DNA are recessed by two bases (4, 11). It is this 3′ end that bonds to the host acceptor DNA (11, 31). Genetic studies show that IN is required for recessing the 3′ ends (4, 31). IN is also involved in the integration of the 3′-recessed intermediate into a target DNA (11, 12). Thus, one can envision that during viral infection reverse transcription produces a linear, blunt-ended viral DNA which is then trimmed by IN at the 3′ end to produce the recombinogenic form of viral DNA.

It is important to note that the reactions described above, when occurring under natural conditions, probably take place within a complex of viral proteins and nucleic acid (3). The approach now being taken by these investigators with respect to integration of MuLV is to isolate the complexes and dissect them (2, 3, 12). This approach is not currently practical with HIV due to difficulties in safely handling large volumes of infectious material. Instead, we are taking a reconstruction approach in which we will first obtain purified IN using recombinant DNA technologies. Using the purified IN protein and a variety of DNA substrates, we are attempting to establish a simple biochemical assay for their interaction, with the ultimate goal of using the assay to screen compounds for anti-integrase activity. For this purpose, we need large quantities of purified IN protein and a suitable DNA substrate. Our progress toward the development of both these reagents is discussed in the next section.

BODY

A. EXPRESSION OF IN PROTEIN

IN is synthesized as a polyprotein; it is cleaved from the 3′ end of the *gag-pol* precursor by the HIV protease encoded by the 5′ end of the *pol* gene. Thus, two general approaches can be used for IN expression in heterologous systems: cloning of the integrase-coding region only, or cloning of all or part of the *pol* polyprotein. The former strategy requires the insertion of an initiator methionine residue at the amino terminus of the protein and therefore results in an altered protein, relative to the native IN. The second strategy produces authentic IN protein but requires the activity of HIV protease, either *in vivo* or *in vitro*.

A1. Expression of IN in baculovirus.

We have expressed IN using both approaches in the baculovirus expression system. To use this system, one constructs an *E. coli* plasmid in which the gene of interest is cloned downstream of the baculovirus polyhedrin promotor. The polyhedrin gene is non-essential for viral growth in cell culture, and proteins expressed from its promotor can account for as much as 50% of total cell protein late in infection (23). Plasmid DNA and baculovirus DNA are then co-transfected into Sf9 cells, and recombinant virus are screened for those that carry the gene of interest in place of the polyhedrin gene. We are exploring three types of construct for baculovirus expression of IN to allow us to compare the yield and activity of the expressed protein (Fig. 1). One type of contruct will express only IN protein from an artificial ATG, one will express the complete POL polyprotein fused to a few amino acids of polyhedrin that will proteolytically process IN protein via the action of the co-expressed protease domain of POL, and one will express the RT and IN domains of POL as a polyprotein and will require *in vitro* processing with exogenous HIV protease to release mature IN. We have now expressed IN in two of these constructs: one containing the entire *pol* gene and one in which IN has an added amino-terminal methionine.

The clone containing only IN was constructed using the polymerase chain reaction (PCR), with an *E. coli pol* subclone as template. Oligonucleotide primers were used to introduce an ATG codon immediately 5′ to the beginning of the IN domain and at the same time to add a BamHI site 5′ to the *int* gene and an EcoRI site 3′ to the gene. The amplified BamHI-EcoRI fragment was first cloned into pBluescriptIIKS+, from which single-stranded DNA can be produced, and the insert was completely sequenced. A fully correct *int* gene was then transferred into the polylinker of the baculovirus expression plasmid pVL1393 (24).

Constructions for the *pol* gene expression were isolated in two fusion vectors, pAc360 and pAc436. These vectors differ in the number of polyhedrin amino acids expressed on the recombinant protein. In pAc360, there is a BamHI site at +36 from the polyhedrin gene initiation codon with which the BglII site at nucleotide 2093 of the NY5′ HIV isolate is in frame; a BglII-BamHI fragment, containing HIV sequences 2093–5464 and a small portion of the pUC19 polylinker derived from one of our subclones was used. In pAc436, we used the SmaI site at +8. SmaI results in blunt-ended fragments; following removal of the 5′ overhang of the BglII site (with mung bean nuclease) and ligation, an in-frame fusion deleting one amino acid from the leader of *pol* was created. The fusion region of each construct was sequenced to confirm that the intended plasmid had been isolated.

Recombinant baculovirus for each of these three constructs have been isolated following co-transfection of the plasmid DNA and wild type baculovirus DNA into Sf9 cells. The recombinants were identified by screening for the plaque morphology characteristic of baculovirus lacking a functional polyhedrin gene. This methodology is rapid and efficient compared to hybridization methods, and it will enable us to continue to explore alternative expression constructs at relatively little cost in time and resources.

Expression of IN by *pol* and *int* constructs is illustrated by Western blot in Fig. 2. The *int* construct yields a large amount of IN protein that is visible by Coomassie staining of complete cell extracts; the *pol* construct expresses a lower level IN and RT that are detectable by Western blotting, but not by Coomassie staining (Fig. 3). Expression from the *pol* clone appears to be maximal about 48 hr after infection; at this time, the accumulation of protein from the *int* clone is still increasing. The lower level of IN expression from the *pol* clone appears to be due to premature cell death and lysis, compared to cells infected with the recombinant *int* virus. Therefore, the *int* clone will be used, at least initially, in developing purification procedures and assays. However, as this protein does carry the additional methionine, it will be compared to native IN from the *pol* clone to detect any differences between the native and recombinant forms.

We are also constructing a *pol* clone containing only RT and IN, which should express an RT-IN polyprotein that remains uncleaved (Fig. 1). The RT portion of the molecule may serve to solubilize the IN moiety, aiding in purification; as a final step, IN would be cleaved from the purified polyprotein using HIV protease, which we have obtained from the NIH AIDS repository. This method would generate a fully authentic IN protein, hopefully without the attendant loss of yield due to expression of the HIV protease *in vivo*. The uncleaved RT-IN polyprotein may also find uses in experiments designed to probe the interactions of these two proteins in the viral core and in the integration reaction. For this construction, we have cloned an 822 bp PstI fragment from a *pol* subclone in pUC19 into pBluescriptIIKS+. The PstI fragment contains the protease gene and the 5′ end of RT. Using Kunkel mutagenesis, an ATG was introduced adjacent to the first codon of RT, and a BamHI site was introduced 15 bp upstream from the ATG. The 307 bp BamHI-PstI fragment, containing only the 5′ end of RT including the ATG, will be recloned into the baculovirus vector pVL1393. The 2.6 kb PstI-BamHI fragment containing the remainder of RT and all of IN, will be cloned downstream, regenerating a complete RT-IN construct with an artificial ATG initiation codon. A recombinant baculovirus expressing this gene will then be isolated and characterized using the methods worked out for our other constructs.

A2. Expression of IN in E. coli.

Our initial strategy was to express IN from the entire *pol* gene so the viral protease could properly process IN from the POL polyprotein. A 3.4 kb AvrII fragment corresponding to nt 2011–5431 of the NY5′ HIV genome was inserted in the XbaI site of the pUC18 multiple cloning site. In one orientation, the insert was subjected to further modification by the removal of 91 bp between the BamHI site of the pUC18 multiple cloning site and a BglII site located at the seventh codon of the *pol* open reading frame. This brings *pol* into frame with *lac*. (This is the same fragment used by Farmerie *et al.* [10]). In the other orientation, the AvrII insert served as a convenient source of a BglII–BamHI fragment containing *pol* for further cloning. The latter fragment was cloned into pIBI20, a pUC-based vector, and into pET3b, a T7-promoter, translational fusion vector developed by W.F. Studier (29). The nucleotide sequences for these constructs were verified. Several different

E. coli hosts were examined for IN expression using Western blotting with a human serum that produces a strong signal at 34kD with viral lysates. In no case did we detect a specific new band corresponding to IN. Background was quite high for all clones carrying viral sequences, including the AvrII insert which was used as an out-of-frame control. This probably reflects internal initation within the viral sequences, since the HIV integrase gene contains a number of potential internal translational initiation sites that may be recognized efficiently by *E. coli* ribosomes. One of these apparently results in the production of a 15 kd fragment of the integrase protein and also greatly reduces overall yields (17). We removed this site by mutagenesis according to the procedure of Kunkel (20) using a mutagenic oligonucleotide hybridizing to nt 4662–4687 of NY5′. Even this construct failed to express IN.

We have now obtained from other investigators clones expressing IN in *E. coli*. Dr. S. Goff (Columbia University) has provided us with p22K56, a *trp* translational fusion vector that produces a fusion protein consisting of 320 residues of *trpE*, 10-14 residues corresponding to the C-terminus of RT, and IN. This protein is expressed at a high level and represents a good source of protein for the preparation of antibodies (Fig. 4). We will also continue to utilize it, in parallel with the IN produced in baculovirus, for developing *in vitro* assays of IN activity. Since the fusion protein contains the RT-IN junction, it should be possible to separate IN from the *trpE*-RT portions by cleavage with HIV protease, if necessary.

A second IN clone was obtained by our collaborators at Sterling Research Group. IN was cloned behind the T7 promoter of the vector pET3 and an initiator methionine was inserted 5′ to the first codon of IN, using a PCR method analagous to that described above for our baculovirus IN clone. We are continuing to study the *E. coli* expression of IN, since such a system may be useful in developing *in vivo* (*E. coli*) assays for IN which may be better suited to drug screening programs. An *E. coli* expression system is also required for future biochemical studies utilizing IN mutants.

B. PURIFICATION OF IN PROTEIN

Preliminary biochemical analysis of IN produced by either type of construct indicated that it is insoluble in low salt buffers. We have found that it is easily solubilized in 0.8 M NaCl (Fig. 5) and a preliminary experiment suggests that IN is also soluble in pH 10 buffers (not shown). There are thus two approaches toward purification. One is to use the insolubility of IN as the initial purification step, i.e. to recover IN from the insoluble pellet following lysis of cells in low salt buffer. The second approach is to lyse cells under conditions that will prevent the precipitation of IN. The first method provides a substantial initial purification step, but could result in inactivation of the protein.

As we begin to assay the recombinant protein and compare it to authentic IN from virus and to the authentic IN from the *pol* construct, we may find that the addition of the N-terminal methionine has a negative impact on the activity of IN. If this proves to be the case, the expression of RT-IN polyprotein and its *in vitro* processing with HIV protease may prove to be the approach of choice. This fusion protein may also have an advantage in terms of the solubility of the expressed protein.

C. CARBOXY TERMINAL PROCESSING OF IN.

The location of the C-terminus of mature integrase is not known with certainty. Because carboxy-terminal processing of the avian integrase is known to occur (1) (18), it

was important to determine whether the HIV integrase is similarly processed. If so, it would have been necessary to introduce a termination codon into the sequence at the appropriate position.

We now have conclusive evidence that integrase is not processed at the carboxy terminus. Antibodies were raised against 2 peptides corresponding to amino acids 4–15 ("N") and 277–288 ("C") of the ARV-2 integrase. The C-terminal peptide C corresponds to the final 12 amino acids of the *pol* open reading frame. Both peptides are highly conserved among different isolates of HIV-1. The antisera were tested by Western blot analysis of viral lysates using the commerically available 3B strain (whose *pol* sequence is not known). Serum raised against the N-terminal and C-terminal peptides reacted with a band of the same size in Western blots of viral lysates (Fig. 6), a band which is also present in positive human sera from AIDS patients. The specificity of the reactions was demonstrated by competition experiments in which the peptide antigens were included during the reaction with antibody: the homologous, but not the heterologous, peptide quantitatively competed out antibody binding. Reactivity to the same band by both anti-carboxy and anti-amino terminal peptides strongly suggests a single, unprocessed form of the viral integrase.

E. DNA SUBSTRATES

E1. Cloned circle junctions

Retroviral DNA is found in three forms in infected cells. Linear DNA, containing a complete LTR at each end, is the product of the reverse transcriptase reaction. Circular DNA, containing either one LTR, or two tandem, directly repeated LTRs, are also present. Despite earlier *in vivo* data implicating the circular, 2-LTR form as the direct precursor for integration (27), it is now generally believed that linear viral DNA is the substrate for integration (3, 11). In order to obtain a source of target DNA for biochemical studies, we have prepared a substrate, using recombinant DNA techniques, that provides us with both circular and linear forms. We have cloned a 248-bp fragment corresponding to the region where the 2 LTRs are joined in circular viral DNA molecules by PCR amplification from infected cells. Template DNA was prepared by the Hirt extraction method (16) from cells infected with pNL4-3. Using the sequences of the NY5′–LAV hybrid LTR and the ARV2 LTR as a guide, oligonucleotide probes 26 and 27 nucleotides in length, hybridizing to nt 594–620 of U5 and 136–109 of U3 (Fig. 7) were purchased. The U5 primer contains a natural HindIII site; a BamHI site was introduced into the U3 primer by altering two nucleotides. This pair of oligonucleotides was used to amplify a fragment 257 bp in length across the circle junction. A small number of discrete fragments, including one of the expected size, were amplified, cleaved with HindIII and BamHI, and cloned into M13mp18. We have obtained the nucleotide sequence of several clones of the circle junction fragment. They are identical to one another and agree with the predicted sequence of the circle junction. The BamHI-HindIII fragment carrying the circle junction has also been subcloned to pUC18, from which large amounts of DNA are more easily obtained than from M13 clones.

In predicting the circle junction sequence, we made the assumption that HIV, like other retroviruses, would have a deletion of 2 bp at each end in the proviral sequence relative to the linear DNA form synthesized by RT. Based on this assumption and on the available sequences, we had predicted that the NY5′ circle junction would contain the 4 additional bp GTAC, forming an RsaI/ScaI (AGTACT) site at the precise position where U3 and U5 are joined. This prediction was borne out; about half the clones obtained had the

expected sequence, indicating that integration does result in the loss of 2 bp at each end. (The remainder of the clones contained insertions or deletions and are discussed below).

As ScaI cleavage yields flush-ended DNA, digestion of our circle junction clones with this enzyme generates a flush-ended linear DNA molecule that has at its ends the precise sequence found in the presumed integration substrate, linear viral DNA. Although the circular form is most likely not a productive intermediate for viral integration, it may be a useful substrate for analysis of interactions with integrase, since endonucleolytic and topoisomerase activities can be measured easily with a circular substrate.

E2. Synthetic substrates.

The initial reaction catalyzed by IN is now believed to be the removal of 2 bp from the 3′ ends of linear viral DNA (4, 11). We have prepared a synthetic substrate for assaying this activity *in vitro*. Two complementary oligonucleotides corresponding to the last 30 bp at the 3′ end of the LTR, plus some additional bases for manipulations, have been made (Fig. 8). When hybridized, one end of the duplex is flush; the other end has a 10 bp, 5′ extension. The shorter oligo will be extended with Klenow, using a varied choice of labeled nucleotides. The now double-stranded, labeled region contains a unique ScaI/RsaI site; when cleaved, a flush-ended 32–mer and a 7-mer are produced. The 7–mer will be labeled in at least one position internally and so the completeness of the restriction digestion can be monitored. Depending on which labeled nucleotides are used in the Klenow reaction, the 3′ end can be differentially labeled in a choice of positions. For example, ^{3}H–dATP can be used to label the 3rd nucleotide from the end and ^{32}P–dGTP and ^{32}P–dTTP to label the last two positions. This will allow us to monitor the loss of the two 3′ nucleotides directly by measuring acid solubility of the ^{32}P label. At the same time we will be able to determine the extent of nonspecific nuclease activity (loss of the tritium label from the oligomer) as well as to study the chemical nature of the reaction (by analysis of the form of the ^{32}P label – i.e. free nucleotides, dinucleotide, free phosphate or pyrophosphate). A similar assay for ASLV endonuclease was recently described (19), which implies that this approach should be successful. In the published assay only the 5′ end of the oligonucleotide was labeled and the length of the fragment remaining after reaction with endonuclease was assayed by fractionation on a sequencing gel. Our assay has the advantage of allowing us to monitor the events at the 3′ end directly and more simply, by an assay that can be relatively easily automated.

E3. Defective circle junctions

In the course of analyzing the PCR products used to obtain the circle junction clone, we found that about half of the cloned circle junctions have abberrant sequences (Fig. 9). Five clones contain deletions: one is a 1 bp deletion in U5 precisely at the circle junction, one has a 113 bp deletion in U3 starting at the circle junction, and three other deletions span the circle junction. These deletions could have arisen during viral replication via incomplete reverse transcription, via cellular exonucleolytic activity, or possibly via an extended IN-catalyzed nuclease activity. Two clones containing insertions precisely at the circle junction were also obtained. One contains a 15 bp insertion, corresponding to the tRNA (primer) binding site (PBS) at the circle junction. This may be a result of reverse transcriptase extending too far during synthesis or of the failure of the RNaseH activity to remove tRNA. The other contained a 340 bp insertion of *gag* sequences, which may have arisen by integration of one viral DNA into another, as shown in Fig. 10.

To ensure that PCR was not preferentially amplifying mutant circle juctions and thereby inflating our estimate of their relative abundance, we performed several reconstruction amplifications with clones containing the wild type circle junction, the 73 bp deletion at the circle junction, and the 340 bp insertion at the circle junction. DNA from the human cell line H9 (4 ug) was mixed with 60 pg of cloned DNA in which the wild type and 73 bp deletion were in ratios of 10:1; 1:1, or 1:10. The mixtures were amplified and sampled after 25, 30, and 35 cycles. In all cases, the PCR products were roughly proportional to the concentration of template DNA clones. Proportional results were also obtained when all three clones were mixed in equal ratios or in 10:10:1 ratios. Thus it is likely that our results do in fact reflect a high proportion of mutant circle junction-containing DNA in infected cells. Similar results were obtained using H9 cells infected with the molecular clone pNL4-3 and with the non-clonal viral isolate ARV-2. We are now planning to extend this analysis to patients' samples to determine if there is a correlation of defective circle formation with the stage of disease or with AZT treatment. This work will be done in collaboration with Ltc. Chet Roberts at WRAIR and Dr. Francine McCuthchan at the Henry M. Jackson Foundation for the Advancement of Military Medicine.

CONCLUSIONS

We have made substantial progress toward production of the two reagents necessary for a study of the biochemical properties of HIV integration. Our multi-pronged approach to the production of recombinant IN has now resulted in at least one clone producing large quantities of IN and several others producing moderate amounts of protein; in addition, we have the option of producing IN as a fusion protein. Protein produced in *E. coli* or eukaryotic cells is available. We believe the major obstacle to purification of IN, namely its insolubility, has been solved, and we are now starting to assess different methods for purification.

We also have constructed several potential DNA substrates. We began this study expecting that the circular, 2 LTR form of viral DNA was the precursor for the integration reaction. Studies with MuLV implicating the linear form (discussed in section 5 above), published during the past year, led us to design linear DNA substrates as well. As soon as an IN preparation free of non-specific nucleases is available, we will begin the development of assays utilizing different DNA substrates and different potential activities of IN, as described in detail in our original proposal. In addition to the 3′ ressection assay described above, these activies include DNA binding, DNA wrapping, topoisomerase and specific nuclease functions. Although none of these involve a complete integration reaction, all have the potential to be automated for screening of inhibitors.

The high proportion of mutant circle junctions in HIV-infected cells was an unexpected result. These mutant circles are certain to be the result of a dead-end pathway. Thus, if an anti-integrase agent increases the fraction of these circles, it could act physiologically as an antiviral "poison" even though biochemically it might only reversibly inactivate IN.

REFERENCES

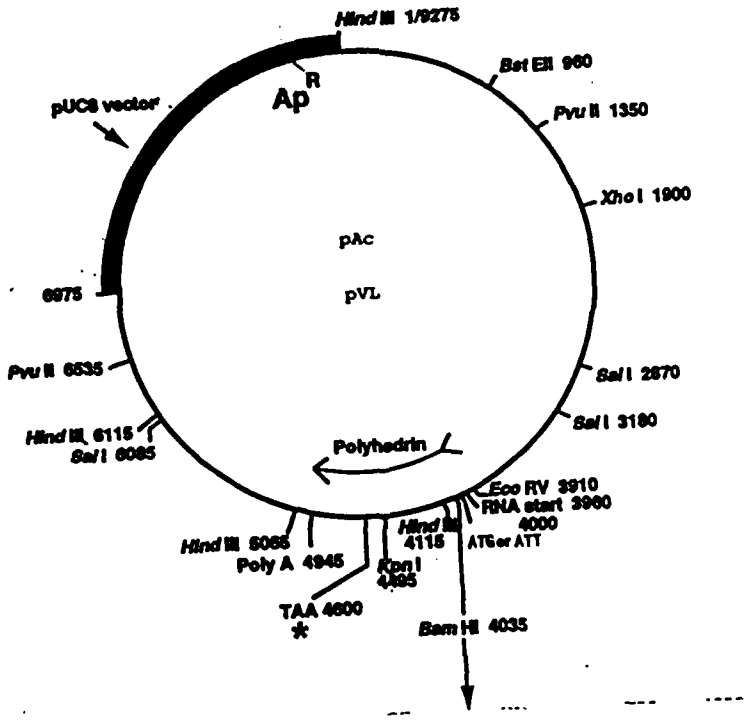
1. Alexander, F., J. Leis, D. A. Soltis, R. M. Crowl, W. Danho, M. S. Poonian, Y. C. E. Pan, and A. M. Skalka. (1987) J. Virol. 61:534-542.
2. Bowerman, B., P. Brown, J. M. Bishop, and H. Varmus. (1989) Genes and Development 3:469-478.
3. Brown, P. O., B. Bowerman, H. E. Varmus, and J. M. Bishop. (1987) Cell 49:347-356.
4. Brown, P., B. Bowerman, H. Varmus, and J. M. Bishop. (1989) Proc. Natl. Acad. Sci. U.S.A. 86:2525-2529.
5. Cobrinik, D., R. Katz, R. Terry, A. M. Skalka, and J. Leis. (1987) Journal of Virology 61:1999-2008.
6. Colicelli, J., and S. P. Goff. (1985) Cell 42:573-580.
7. Donehower, L. A., and H. E. Varmus. (1984) Proc. Natl. Acad. Sci. USA 81:6461-6465.
8. Duyk, G., M. Longiaru, D. Cobrinik, R. Kowal, P. Dehaseth, A. Skalka, and J. Leis. (1985) J. Virol. 56:589-599.
9. Ellis, J., and A. Bernstein. (1989) J. Virol. 63:2844-2846.
10. Farmerie, W., D. Loeb, N. Casavant, C. H. III, M. Edgell, and R. Swanstrom. (1987) Science 236:305-308.
11. Fujiwara, T., and K. Mizuuchi. (1988) Cell 54:497-504.
12. Fujiwara, T., and R. Craigie. (1989) Proc. Natl. Acad. Sci. U.S.A. 86:3065-3069.
13. Golomb, M., and D. P. Grandgenett. (1979) J. Biol. Chem. 254:1606-1613.
14. Grandgenett, D. P., and A. C. Vora. (1985) Nucl. Acids Res. 13:6205-6221.
15. Grandgenett, D. P., A. C. Vora, R. Swanstrom, and J. C. Olsen. (1986) J. Virol. 58:970-974.
16. Hirt, B. (1967) J. Mol. Biol. 26:362-371.
17. Hizi, A., and S. Hughes. (1988) Virology 167:634-638.
18. Katz, R. A., and A. M. Skalka. (1987) J. Virol. 62:528-533.
19. Katzman, M., R. A. Katz, A. M. Skalka, and J. Leis. (1989) J. Virology 63:5319-5327.
20. Kunkel, T. A., J. D. Roberts, and R. A. Zakour. (1987) Meth. Enzymol. 154:367-383.
21. Leis, J., G. Duyk, S. Johnson, M. Longiaru, and A. Skalka. (1983) Journal of Virology 45:727-739.
22. Lobel, L., J. Murphy, and S. Goff. (1989) J. Virol. 63:2629-2637.
23. Luckow, V., and M. Summers. (1988) Biotechnology 6:47-54.
24. Luckow, V., and M. Summers. (1989) Virology 170:31-39.
25. Misra, T. K., D. P. Grandgenett, and J. T. Parsons. (1982) Journal of Virology 44:330-343.
26. Panganiban, A. T., and H. M. Temin. (1983) Nature 306:155-160.
27. Panganiban, A. T., and H. M. Temin. (1984) Cell 36:673-679.

28. Quinn, T., and D. Grandgenett. (1988) J. Virol. 62:2307-2312 .

29. Rosenberg, A., B. Lade, D. Chui, S. Lin, J. Dunn, and W. Studier. (1987) Gene 56:125-135.

30. Schwartzberg, P., J. Colicelli, and S. P. Goff. (1984) Cell 37:1043-1052.

31. Tanese, N., M. Roth, H. Epstein, and S. P. Goff. (1989) Virology 170:378-384.



STRATEGIES FOR INTEGRASE EXPRESSION

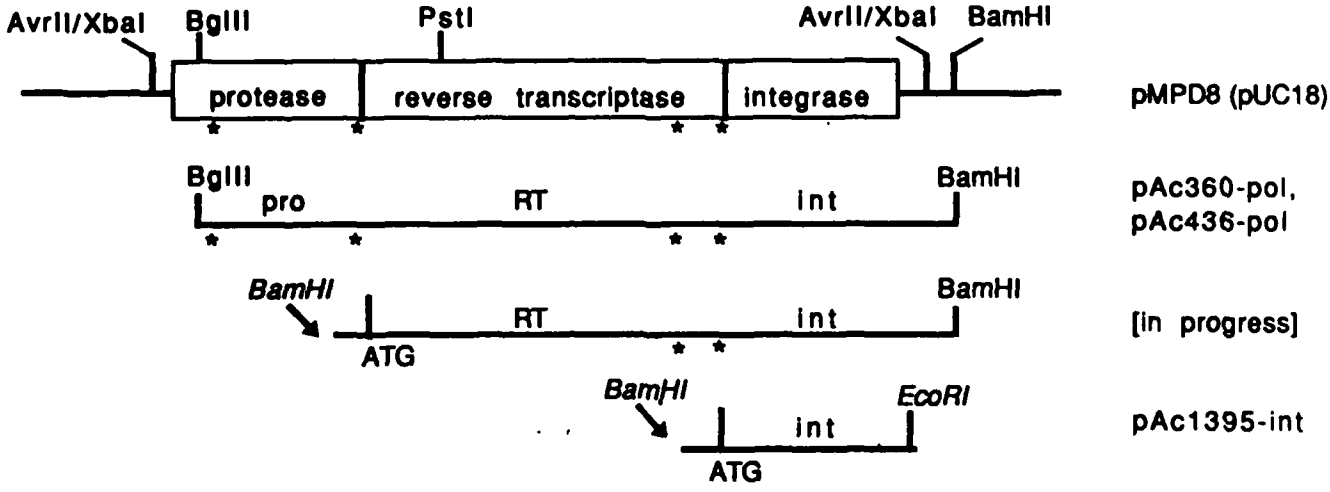


Figure 1. Construction of baculovirus clones expressing IN. The upper half shows a general map of the baculovirus vectors. The lower half shows the strategy for expression of IN. The source of *pol* sequences was pMPD8, an *E. coli* subclone which has the AvrII fragment nt 2018–5221 from NY5′ inserted in the XbaI site of pUC18. Asterisks indicate processing sites for HIV protease. Restriction sites in italics are those created by site-specific mutagenesis.

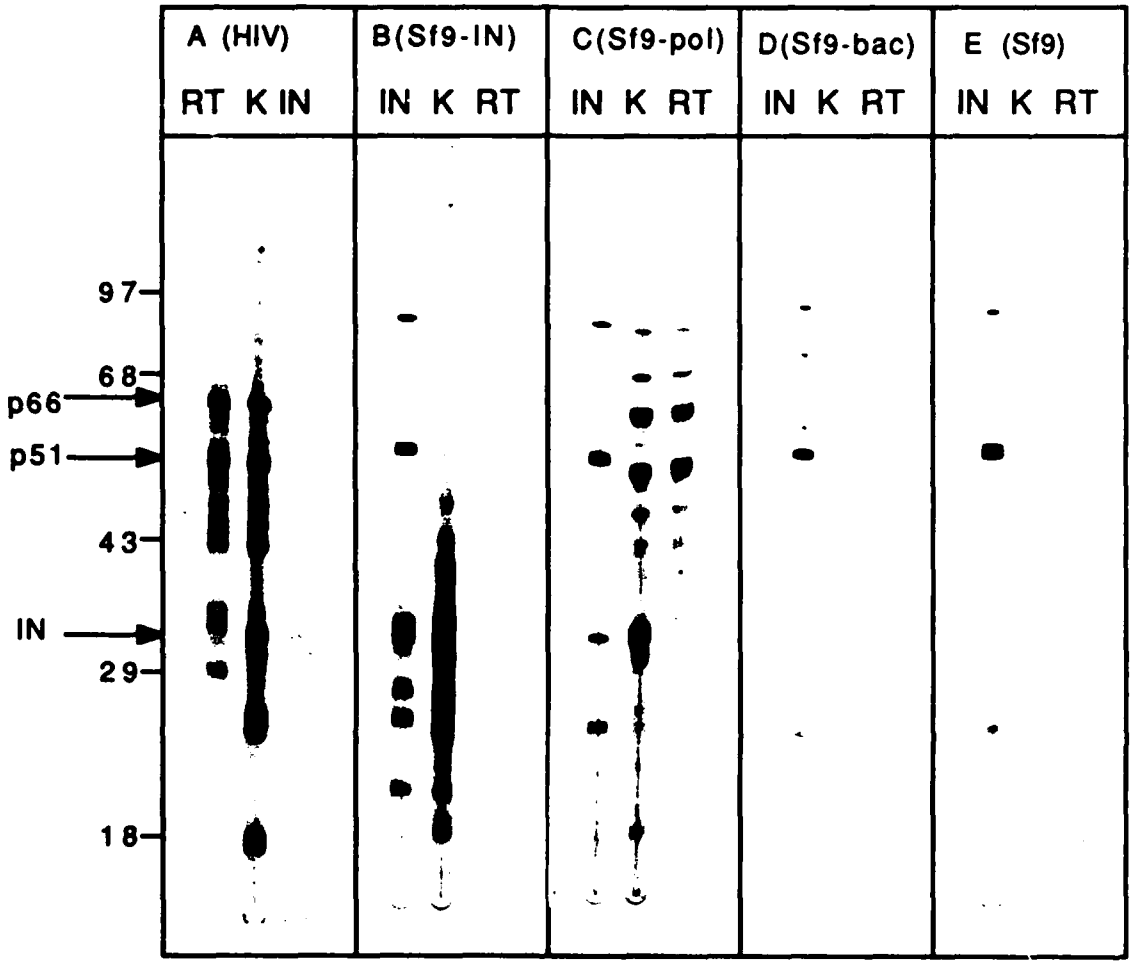


Figure 2.. Western blot analysis of HIV-1 *pol* proteins produced in Sf9 insect cells infected with recombinant baculovirus. Sf9 cells were infected with different recombinant baculoviruses and 42 hr after infection cells were lysed with 1% NP40 in 0.01 M Tris (pH 7.4). Samples were boiled for 10 min in 0.01% bromphenol blue, 10% glyerol 0.1 M Tris (pH 8.0), 1% DTT 1% SDS and then electrophoresed in a 11% polyacrylamide gel. Proteins were transferred to nitrocellulose, and the blots were incubated with serum from and AIDS patient (lane K), a human monoclonal antibody directed agains HIV-RT (lane RT), and a rabbit polyclonal antibody directed agains the N-terminal part of IN (lane IN). Panel A: HIV-1 viral lysate, Panel B: recombinant baculovirus int1993-infected cells (IN-only), Panel C: recombinat baculovirus pol360-infected cells (entire *pol* gene), Panel D: wild type baculovirus-infected cells, and Panel E: uninfected Sf9 cells. Positions of molecular weight standards are indicated on the left. Arrows indicate the positions of IN and RT (p51 and p66).

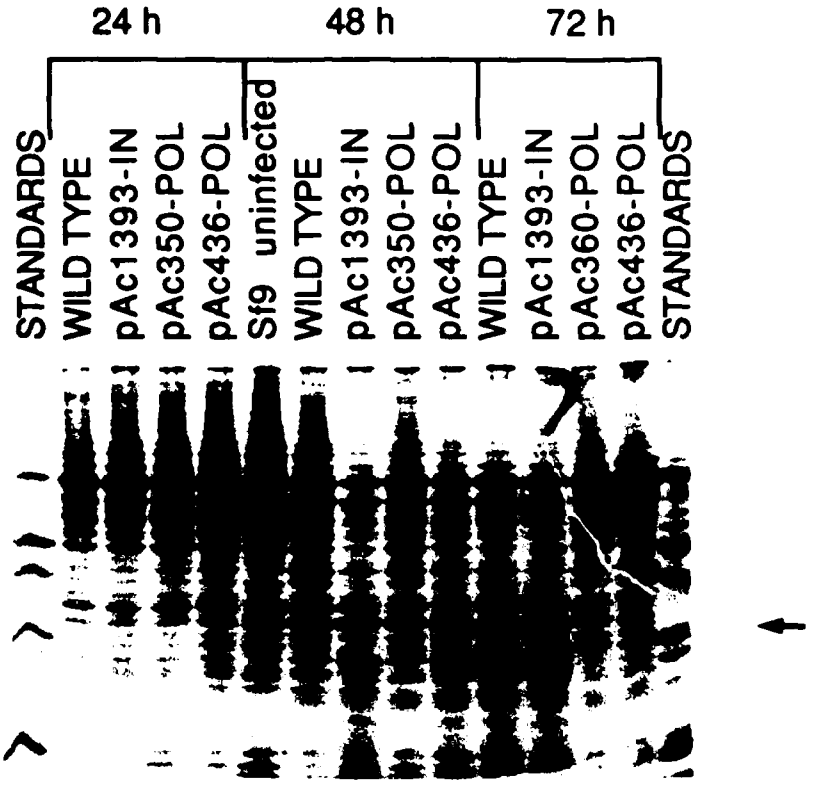


Figure 3. IN expression in baculovirus. This Coomasie-stained gel shows a time course of expression of IN in baculovirus from the pAc1393-IN and pAC360-*pol* clones. IN, marked by the arrow on the right is clearly visible at 48 and 72 hr in the pAc1393-IN clones but is not detectable in the other samples. The strong band in the lanes corresponding to infection with wild-type baculovirus, migrating slightly faster than IN, is the polyhedrin protein.

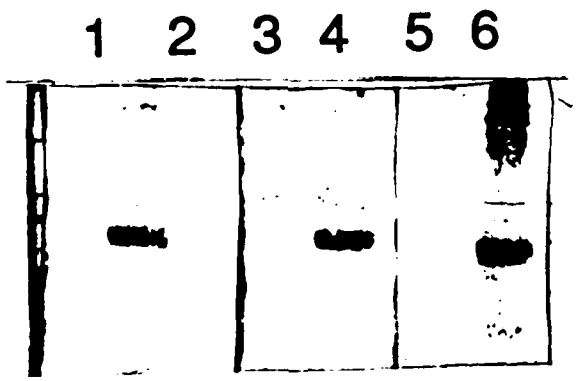


Figure 4. Expression of *trpE*-IN fusion. HB101 cells carrying the plasmid p22K56 were induced with IAA (indole acrylic acid), 5 ug/ml, for 2 hr, lysed by the addition of loading buffer, separated by SDS-PAGE and analyzed by Western blotting. Three antisera were used: lanes 1 and 2, anti-*pol* obtained from C. Debouck at Smith Kline & French; lanes 3 and 4, antiserum raised against a C-terminal IN peptide; lanes 5 and 6, an AIDS patient serum. Lanes 2, 4 and 6 contain p22K56 and alnes 1, 3 and 5 contain the *trpE* vector lacking any insert.

Figure 5. Salt-dependent solubility of IN protein produced in Sf9 cells. Cells were infected with recombinant baculoviruses Ac360-*pol* (lane 1) and Ac1393-IN (lanes 2-6). 42 hr after infection, cells were lysed in 10 mM HEPES pH 7.0, 50 mM NaCl, 0.5mM PMSF, 1% NP-40, 1% Triton X-100, 1% deoxycholate (lanes 1-4) or the same buffer containing 800 mM NaCl (lanes 5 and 6). The lysates were centrifuged for 5 min at 14,000 rpm. Equal volumes of the total lysate (T), pellet (P) or supernatant (S) fractions were boiled for 10 min in Laemmli sample buffer (0.01% bromphenol blue, 10% glycerol, 0.01M Tris pH 8.0, 1% dithiothreitol, 1% SDS) and electrophoresed in an 11% SDS-polyacrylamide gel. Western blotting of the gel was performed using serum from an AIDS patient. **Figure 6.** Lack of carboxy-terminal processing of IN. Western blots of commercial HIV3B lysate reacted with antisera raised against N-terminal (Panel A) and C-terminal (Panel B) peptides. In each set the three lanes represent pre-immune serum, antibody only, and antibody plus the homologous peptide antigen. Controls with heterologous peptides are not shown on this blot.

REACTION OF VIRAL IN WITH
ANTI-N AND ANTI-C TERMINAL ANTIBODIES

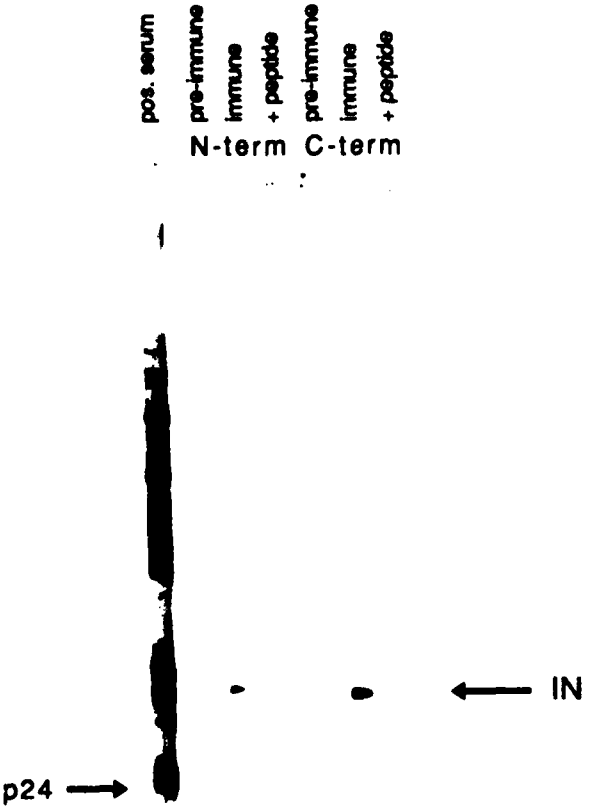


Figure 6. Lack of carboxy-terminal processing of IN. Western blots of commercial HIV3B lysate reacted with antisera raised against N-terminal and C-terminal peptides, as described in the text. In each set the three lanes represent pre-immune serum, antibody only, and antibody plus the homologous peptide antigen. Controls with heterologous peptides are not shown on this blot.

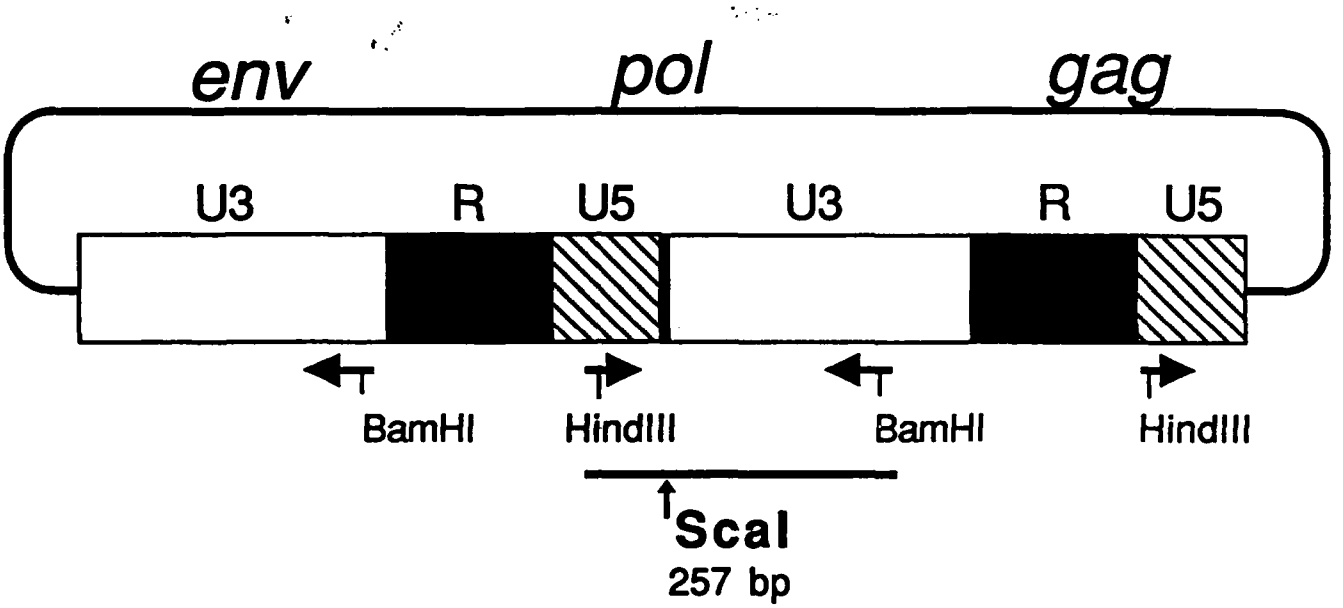


Figure 7. Cloning strategy for circle junctions. Two oligonucleotide primers were prepared: a 26-mer hybridizing to U5 and containing a natural and unique HindIII site, and a 27-mer hybridizing to U3 with two alterations to create a unique BamHI site at the 5′ end. A 257 bp fragment was amplified from H9 cells infected by ARV2 or pNL4-3 (NY5/LAV hybrid). The fragment was cleaved with HindIII and BamHI to generate sticky ends and cloned in M13. Although each primer hybridizes to two locations in the 2-LTR circle, only one of the primer pairs is productive for PCR.

IN ASSAY. WITH SYNTHETIC OLIGONUCLEOTIDE SUBSTRATE

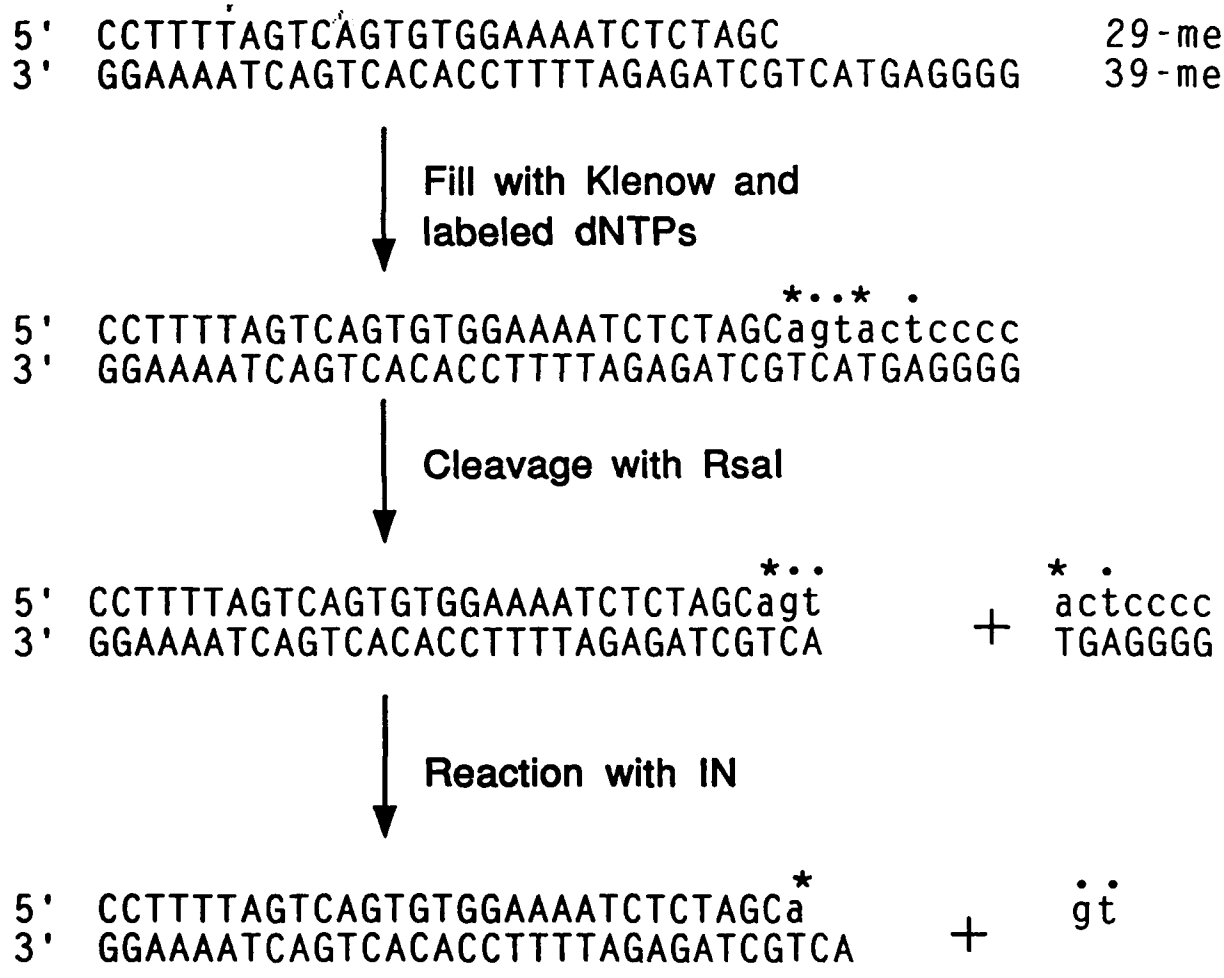


Figure 8 Synthetic substrate for *in vitro* assay of IN protein. The substrate consists of a complementary 29 and 39-mer corresponding to the 3′ end of the HIV LTR. After filling in the 29-mer with Klenow fragment and ^{3}H–dATP (*) and ^{32}P–dGTP (•) and ^{32}P–dTTP (•) and cleavage with RsaI, the flush-ended 32-mer, labeled in three positions, is purified from the 7-mer. After reaction with IN, which is expected to remove two 3′ nucleotides, the oligomer remains labeled with tritium at the 3′ end; the ^{32}P label is associated with a GT dinucleotide as shown, or with free nucleotides or free phosphate.

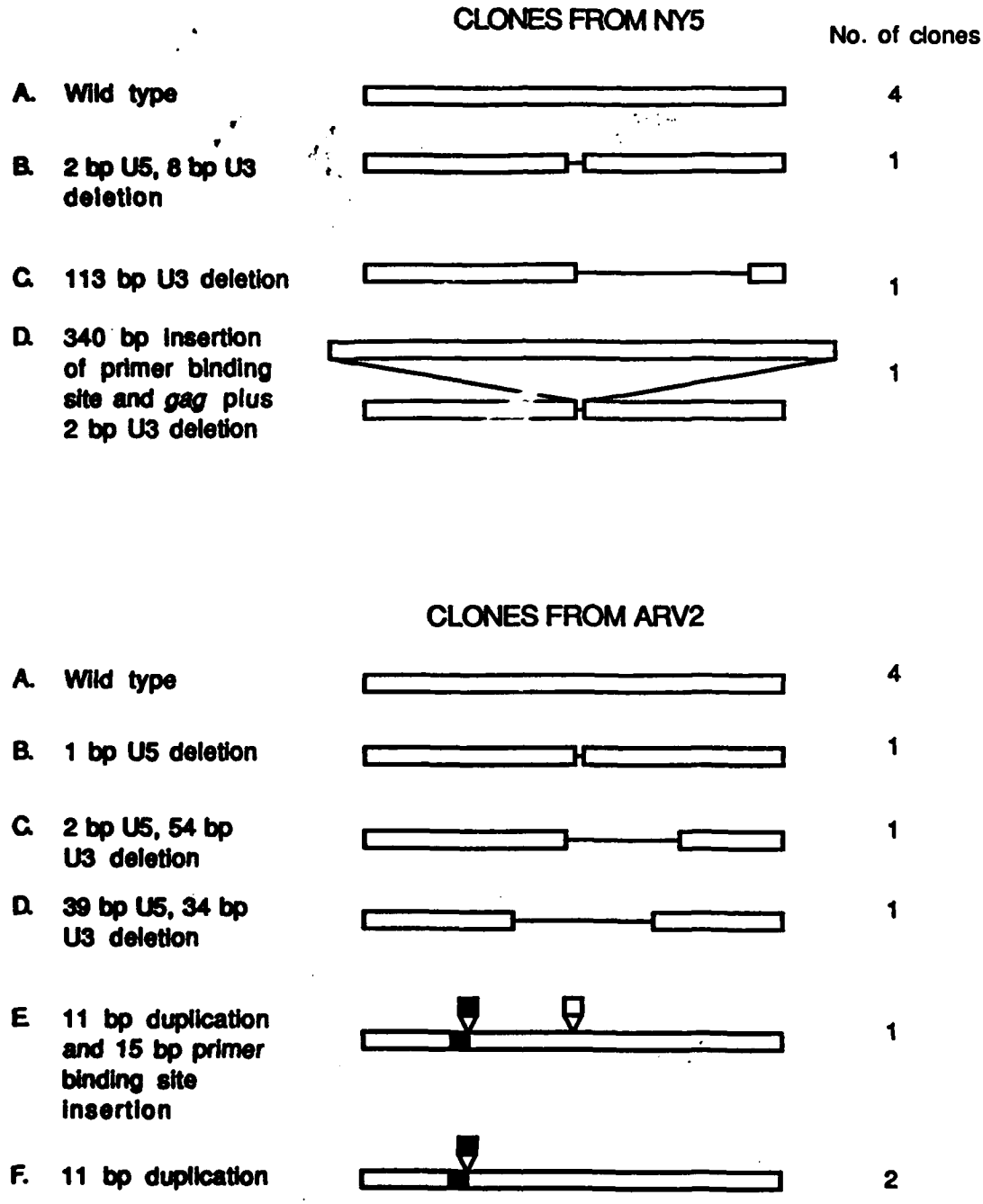


Figure 9. Summary of mutant circle junction clones obtained by PCR amplification from cells infected with HIVNL4-3 (NY5, top) and ARV2 (bottom). Thin lines indicate deletions.

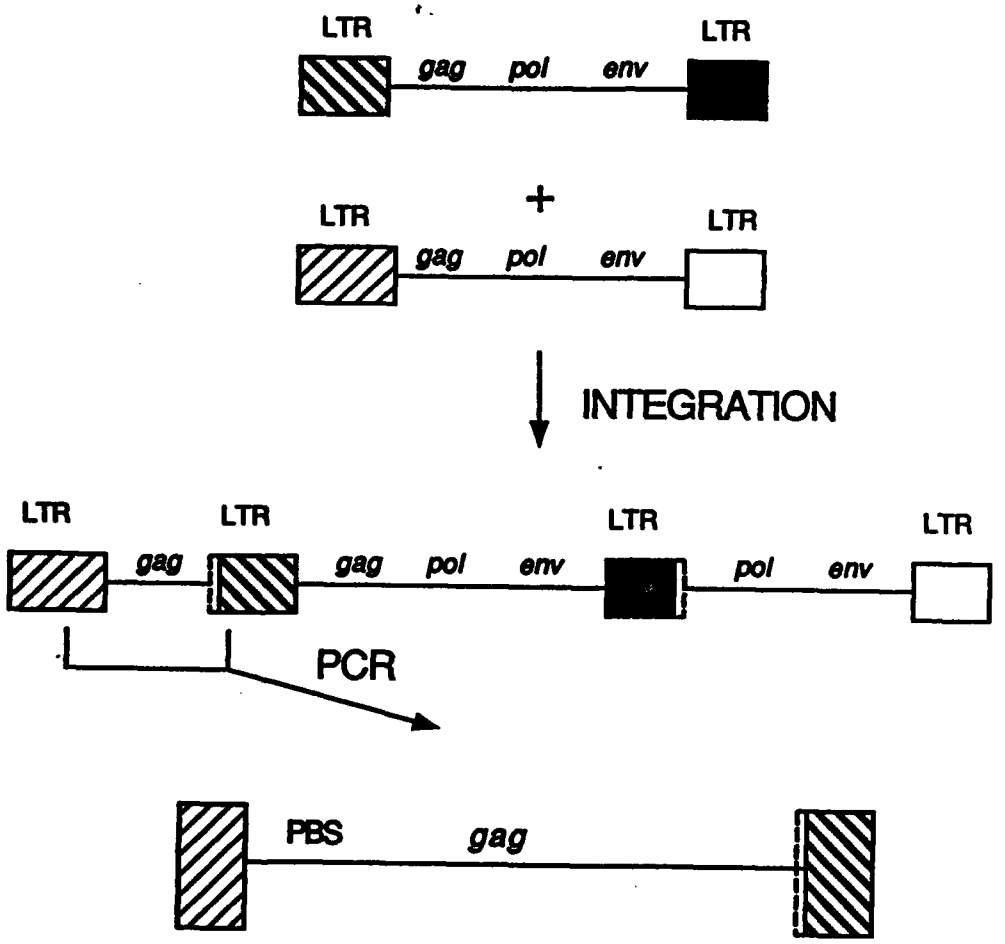


Figure 10. A possible pathway for the generation of a circle junction clone containing a 340 bp insertion of *gag* DNA via integration of one viral molecule into another. The absence of 2 bp from the 5′ end of the U3 DNA in the PCR clone (indicated by the dotted lines) suggests that this configuration arose via a normal integration event, in this case that of one viral DNA molecule integrating into the *gag* gene of a second. PBS, tRNA (primer) binding site.